Efficient robust nonparametric estimation in a semimartingale regression model *

Konev Victor [†]and Pergamenshchikov Serguei[‡]

October 19, 2010

Abstract

The paper considers the problem of robust estimating a periodic function in a continuous time regression model with dependent disturbances given by a general square integrable semimartingale with unknown distribution. An example of such a noise is non-gaussian Ornstein-Uhlenbeck process with the Lévy process subordinator, which is used to model the financial Black-Scholes type markets with jumps. An adaptive model selection procedure, based on the weighted least square estimates, is proposed. Under general moment conditions on the noise distribution, sharp non-asymptotic oracle inequalities for the robust risks have been derived and the robust efficiency of the model selection procedure has been shown.

Keywords: Non-asymptotic estimation; Robust risk; Model selection; Sharp oracle inequality; Asymptotic efficiency.

AMS 2000 Subject Classifications: 62G08, 62G05

*The paper is supported by the RFBR-Grant 09-01-00172-a.

[†]Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia, e-mail: vvkonev@mail.tsu.ru

[‡]Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France and Department of Mathematics and Mechanics,Tomsk State University, Lenin str. 36, 634041 Tomsk, Russia, e-mail: Serge.Pergamenchtchikov@univ-rouen.fr

1 Introduction

Consider a regression model in continuous time

$$\mathrm{d}y_t = S(t)\mathrm{d}t + \mathrm{d}\xi_t\,, \quad 0 \le t \le n\,, \tag{1.1}$$

where S is an unknown 1-periodic $\mathbb{R} \to \mathbb{R}$ function, $S \in \mathcal{L}_2[0,1]$; $(\xi_t)_{t\ge 0}$ is an unobservable semimartingale noise with the values in the Skorokhod space $\mathcal{D}[0,n]$ such that, for any function f from $\mathcal{L}_2[0,n]$, the stochastic integral

$$I_n(f) = \int_0^n f_s \mathrm{d}\xi_s \tag{1.2}$$

is well defined and has the following properties

$$\mathbf{E}_Q I_n(f) = 0 \quad \text{and} \quad \mathbf{E}_Q I_n^2(f) \le \sigma_Q \int_0^n f_s^2 \mathrm{d}s\,. \tag{1.3}$$

Here $\mathbf{E}_Q$ denotes the expectation with respect to the distribution Q in $\mathcal{D}[0,n]$ of the process $(\xi_t)_{0\le t\le n}$, which is assumed to belong to some probability family $\mathcal{Q}_n$ specified below; $\sigma_Q > 0$ is some positive constant depending on the distribution Q.

The problem is to estimate the unknown function S in the model (1.1) on the basis of observations $(y_t)_{0\le t\le n}$.

The class of the disturbances ξ satisfying conditions (1.3) is rather wide and comprises, in particular, the Lévy processes which are used in different important problems (see [4], for details). The models (1.1) with the Lévy's type noise naturally arise (see [18]) in the nonparametric functional statistics problems (see, for example, [8]). Moreover, as is shown in Section 2, Non-Gaussian Ornstein-Uhlenbeck-based models also enter this class. The latter models are successfully used to model the Black-Scholes type financial markets with jumps (see [2], [6] for details and other references).

We define the error of an estimate $\widehat{S}$ (any real-valued function measurable with respect to $\sigma\{y_t\,,\, 0 \le t \le n\}$) for S by its integral quadratic risk

$$\mathcal{R}_Q(\widehat{S}, S) := \mathbf{E}_{Q,S}\,\|\widehat{S} - S\|^2\,, \tag{1.4}$$

where $\mathbf{E}_{Q,S}$ stands for the expectation with respect to the distribution $\mathbf{P}_{Q,S}$ of the process (1.1) with a fixed distribution Q of the noise $(\xi_t)_{0\le t\le n}$ and a given function S; $\|\cdot\|$ is the norm in $\mathbf{L}_2[0,1]$, i.e.

$$\|f\|^2 := \int_0^1 f^2(t)\mathrm{d}t\,. \tag{1.5}$$

Since in our case the noise distribution Q is unknown, it seems natural to measure the quality of an estimate $\widehat{S}$ by the robust risk defined as

$$\mathcal{R}_n^*(\widehat{S}, S) = \sup_{Q\in\mathcal{Q}_n} \mathcal{R}_Q(\widehat{S}, S) \tag{1.6}$$

which assumes taking supremum of the error (1.4) over the whole family of admissible distributions $\mathcal{Q}_n$.

It is natural to treat the stated problem with respect to the quadratic risk from the standpoint of the model selection approach. It will be noted that the origin of this method goes back to early seventies with the pioneering papers by Akaike [1] and Mallows [21] who proposed to introduce penalizing in a log-likelihood type criterion. The further progress has been made by Barron, Birgé and Massart [3], [22], who developed a non-asymptotic model selection method which enables one to derive non-asymptotic oracle inequalities for nonparametric regression models with the i.i.d. gaussian disturbances. An oracle inequality yields the upper bound for the estimate risk via the minimal risk corresponding to a chosen family of estimates. Galtchouk and Pergamenshchikov [9] applied the Barron-Birgé-Massart technic to the problem of estimating a nonparametric drift function in ergodic diffusion processes. Fourdrinier and Pergamenshchikov [7] extended the Barron-Birgé-Massart method to the models with the spherically symmetric dependent observations. They proposed a model selection procedure based on the improved least squares estimates. Lately, the authors [17] applied this method to the nonparametric problem of estimating a periodic function in a model with a gaussian colored noise in continuous time. In all cited papers, the non-asymptotic oracle inequalities have been derived, which enable one to establish the optimal convergence rate for the minimax risks. In addition to the optimal convergence rate, the other important problem is that of the efficiency of adaptive estimation procedures. In order to examine the efficiency property of a procedure one has to obtain the *sharp oracle inequalities*, i.e. such in which the factor at the principal term in the right-hand of the inequality is close to unity.

The first result on sharp inequalities is most likely due to Kneip [15] who studied a gaussian regression model. It will be observed that the derivation of oracle inequalities usually rests upon the fact that the initial model, by applying the Fourier transformation, is reduced to the gaussian model with independent observations. However, such a transform is possible only for gaussian models with independent homogeneous observations or for the inhomogeneous ones

with the known correlation characteristics. This restriction significantly narrows the area of application of the proposed model selection procedures and rules out a broad class of models including, in particular, widely used in econometrics heteroscedastic regression models (see, for example, [14]). For constructing adaptive procedures in the case of inhomogeneous observations one needs to modify the approach to the estimation problem. Galtchouk and Pergamenshchikov [11]-[12] have developed a new estimation method intended for the heteroscedastic regression models in discrete time. The heart of this method is to combine the Barron-Birgé-Massart non-asymptotic penalization method [3] and the Pinsker weighted least square method minimizing the asymptotic risk (see, for example, [23], [24]). This yields a significant improvement in the performance of the procedure (see numerical example in [11]).

The goal of this paper is to develop the robust efficient model selection method for the model (1.1) with dependent disturbances having unknown distribution. We follow the approach, proposed by Galtchouk and Pergamenshchikov in [11], in the construction of the procedure. Unfortunately, their method of obtaining the oracle inequalities is essentially based on the independence of observations and can not be applied here. The paper proposes the new analytical tools which allow one to obtain the sharp non-asymptotic oracle inequalities for robust risks under general conditions on the distribution of the noise in the model (1.1). This method enables us to treat both the cases of dependent and independent observations from the same standpoint, does not assume the knowledge of the noise distribution and leads to the efficient estimation procedure with respect to the risk (1.6). The validity of the conditions imposed on the noise in the equation (1.1) is verified for a non-gaussian Ornstein-Uhlenbeck process (see Section 2).

The rest of the paper is organized as follows. In Section 3 we construct the model selection procedure on the basis of weighted least squares estimates and state the main results in the form of oracle inequalities for the quadratic risk (1.4) and the robust risk (1.6). Here we specify also the set of admissible weight sequences in the model selection procedure. In Section 4 we proof some properties of the stochastic integrals with respect to the non-gaussian Ornstein-Uhlenbeck process (2.1). Section 5 gives the proofs of the main results. In Sections 6, 7 it is shown that the proposed model selection procedure for estimating S in (1.1) is asymptotically efficient with respect to the robust risk

(1.6). In Appendix some auxiliary propositions are given.

2 Non-Gaussian Ornstein-Uhlenbeck process

In this section we consider an important example of the disturbances $(\xi_t)_{t\geq 0}$ in (1.1) given by a non-gaussian Ornstein-Uhlenbeck process with the Lévy subordinator. Such processes are used in the financial Black-Scholes type markets with jumps (see, for example [6] and the references therein). Let the noise process in (1.1) obey the equation

$$\mathrm{d}\xi_t = a\xi_t \mathrm{d}t + \mathrm{d}u_t\,, \quad \xi_0 = 0\,, \tag{2.1}$$

where $a \leq 0$, $u_t = \varrho_1 w_t + \varrho_2 z_t$, ϱ_1 and ϱ_2 are unknown constants, $(w_t)_{t\geq 0}$ is a standard Brownian motion, $(z_t)_{t\geq 0}$ is a compound Poisson process defined as

$$z_t = \sum_{j=1}^{N_t} Y_j\,.$$

Here $(N_t)_{t\geq 0}$ is a standard homogeneous Poisson process with unknown intensity $\lambda > 0$ and $(Y_j)_{j\geq 1}$ is an i.i.d. sequence of random variables with

$$\mathbf{E}Y_j = 0\,, \quad \mathbf{E}Y_j^2 = 1 \quad \text{and} \quad \mathbf{E}Y_j^4 < \infty\,. \tag{2.2}$$

Let $(T)_{k\geq 1}$ denote the arrival times of the process $(N_t)_{t\geq 0}$, that is,

$$T_k = \inf\{t \geq 0\,:\, N_t = k\}\,. \tag{2.3}$$

We assume that the parameters λ, a, ϱ_1 and ϱ_2 satisfy the conditions

$$-\,a_{max} \leq a \leq 0\,, \quad 0 \leq \lambda \leq \lambda_{max}\,, \quad \varrho^*_{min} \leq \varrho^* \leq \varrho^*_{max}\,, \tag{2.4}$$

where $\varrho^* = \varrho_1^2 + \lambda\varrho_2^2$. Let $\mathcal{Q}_n$ denote the family of all distributions of process (2.1) on $\mathcal{D}[0,n]$ with the parameters a, λ, ϱ_1 and ϱ_2 satisfying the conditions (2.4) with fixed bounds $\lambda_{max} > 0$, $a_{max} > 0$, $\varrho^*_{min} > 0$ and $\varrho^*_{max} > 0$.

3 Model selection

This Section gives the construction of a model selection procedure for estimating a function S in (1.1) on the basis of weighted least square estimates and states the main results.

For estimating the unknown function S in the model (1.1), we apply its Fourier expansion in the trigonometric basis $(\phi_j)_{j\ge 1}$ in $\mathcal{L}_2[0,1]$ defined as

$$\phi_1 = 1\,, \quad \phi_j(x) = \sqrt{2}\,Tr_j(2\pi[j/2]x)\,,\ j \ge 2\,, \tag{3.1}$$

where the function $Tr_j(x) = \cos(x)$ for even j and $Tr_j(x) = \sin(x)$ for odd j; $[x]$ denotes the integer part of x. The corresponding Fourier coefficients

$$\theta_j = (S, \phi_j) = \int_0^1 S(t)\,\phi_j(t)\,\mathrm{d}t \tag{3.2}$$

can be estimated as

$$\widehat{\theta}_{j,n} = \frac{1}{n}\int_0^n \phi_j(t)\,\mathrm{d}y_t\,. \tag{3.3}$$

In view of (1.1), we obtain

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n}\,, \quad \xi_{j,n} = \frac{1}{\sqrt{n}} I_n(\phi_j) \tag{3.4}$$

where $I_n(\phi_j)$ is given in (1.2).
For any sequence $x = (x_j)_{j\ge 1}$, we set

$$|x|^2 = \sum_{j=1}^{\infty} x_j^2 \quad \text{and} \quad \#(x) = \sum_{j=1}^{\infty} \mathbf{1}_{\{|x_j|>0\}}\,. \tag{3.5}$$

Now we impose some additional conditions on the distribution of the noise $(\xi_t)_{t\ge 0}$ in (1.1).

$\mathbf{C}_1$) *There exists a positive constant $\varsigma_Q > 0$ such that for any $n \ge 1$*

$$\mathbf{L}_{1,n}(Q) = \sup_{x\in\mathcal{H}\,,\,\#(x)\le n} \left| \sum_{j=1}^{\infty} x_j\,\left(\mathbf{E}_Q\,\xi_{j,n}^2 - \varsigma_Q\right) \right| < \infty\,,$$

where $\mathcal{H} = [-1,1]^\infty$.

$\mathbf{C}_2$) *Assume that for all* $n \geq 1$

$$\mathbf{L}_{2,n}(Q) = \sup_{|x|\leq 1\,,\,\#(x)\leq n} \mathbf{E}_Q \left(\sum_{j=1}^{\infty} x_j \, (\xi_{j,n}^2 - \mathbf{E}_Q \xi_{j,n}^2) \right)^2 < \infty \,.$$

As is shown in the proof of Theorem 3.2 in Section 5 , both Conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) hold for the process (2.1). Further we define a class of weighted least squares estimates for $S(t)$ as

$$\widehat{S}_\gamma = \sum_{j=1}^{\infty} \gamma(j) \widehat{\theta}_{j,n} \phi_j \,, \tag{3.6}$$

where $\gamma = (\gamma(j))_{j\geq 1}$ is a sequence of weight coefficients such that

$$0 \leq \gamma(j) \leq 1 \quad \text{and} \quad 0 < \#(\gamma) \leq n \,. \tag{3.7}$$

Let Γ denote a finite set of weight sequences $\gamma = (\gamma(j))_{j\geq 1}$ with these properties, $\nu = \operatorname{card}(\Gamma)$ be its cardinal number and

$$\mu = \max_{\gamma\in\Gamma} \#(\gamma) \,. \tag{3.8}$$

The model selection procedure for the unknown function S in (1.1) will be constructed on the basis of a family of estimates $(\widehat{S}_\gamma)_{\gamma\in\Gamma}$. The choice of a specific set of weight sequences Γ is discussed at the end of this section. In order to find a proper weight sequence γ in the set Γ one needs to specify a cost function. When choosing an appropriate cost function one can use the following argument. The empirical squared error

$$\operatorname{Err}_n(\gamma) = \|\widehat{S}_\gamma - S\|^2$$

can be written as

$$\operatorname{Err}_n(\gamma) \,=\, \sum_{j=1}^{\infty} \gamma^2(j) \widehat{\theta}_{j,n}^2 \,-\, 2 \sum_{j=1}^{\infty} \gamma(j) \widehat{\theta}_{j,n} \, \theta_j \,+\, \sum_{j=1}^{\infty} \theta_j^2 \,. \tag{3.9}$$

Since the Fourier coefficients $(\theta_j)_{j\geq 1}$ are unknown, the weight coefficients $(\gamma_j)_{j\geq 1}$ can not be determined by minimizing this quantity.

To circumvent this difficulty one needs to replace the terms $\widehat{\theta}_{j,n}\,\theta_j$ by some their estimators $\widetilde{\theta}_{j,n}$. We set

$$\widetilde{\theta}_{j,n} = \widehat{\theta}^2_{j,n} - \frac{\widehat{\sigma}_n}{n} \tag{3.10}$$

where $\widehat{\sigma}_n$ is some estimator for the quantity ς_Q in the condition $\mathbf{C}_1$).

For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\gamma) \,=\, \sum_{j=1}^{\infty} \gamma^2(j)\widehat{\theta}^2_{j,n} \,-2\,\sum_{j=1}^{\infty} \gamma(j)\,\widetilde{\theta}_{j,n} \,+\, \rho\,\widehat{P}_n(\gamma) \tag{3.11}$$

where ρ is some positive constant, $\widehat{P}(\gamma)$ is the penalty term defined as

$$\widehat{P}_n(\gamma) = \frac{\widehat{\sigma}_n\,|\gamma|^2}{n}\,. \tag{3.12}$$

In the case, when the value of σ in $\mathbf{C}_1$) is known, one can take $\widehat{\sigma}_n = \varsigma_Q$ and

$$P_n(\gamma) = \frac{\varsigma_Q\,|\gamma|^2}{n}\,. \tag{3.13}$$

Substituting the weight coefficients, minimizing the cost function

$$\widehat{\gamma} = \mathrm{argmin}_{\gamma\in\Gamma}\, J_n(\gamma)\,, \tag{3.14}$$

in (3.6) leads to the model selection procedure

$$\widehat{S}_* = \widehat{S}_{\widehat{\gamma}}\,. \tag{3.15}$$

It will be noted that $\widehat{\gamma}$ exists, since Γ is a finite set. If the minimizing sequence in (3.14) $\widehat{\gamma}$ is not unique, one can take any minimizer.

Theorem 3.1. *Let $\mathcal{Q}_n$ be a family of the distributions Q on $\mathcal{D}[0,n]$ such that the conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) hold for each Q from $\mathcal{Q}_n$. Then for any $n\ge 1$ and $0<\rho<1/3$, the estimator (3.15), for each $Q\in\mathcal{Q}_n$, satisfies the oracle inequality*

$$\mathcal{R}_Q(\widehat{S}_*,S) \,\le\, \frac{1+3\rho-2\rho^2}{1-3\rho}\,\min_{\gamma\in\Gamma}\mathcal{R}_Q(\widehat{S}_\gamma,S) + \frac{1}{n}\,\mathcal{B}_Q(n,\rho) \tag{3.16}$$

where the risk $\mathcal{R}_Q(\cdot,S)$ is defined in (1.4),

$$\mathcal{B}_Q(n,\rho) = \Psi_Q(n,\rho) + \frac{6\mu\,\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \varsigma_Q|}{1-3\rho}$$

and

$$\Psi_Q(n,\rho) = \frac{2\varsigma_Q \sigma_Q \nu + 4\varsigma_Q \mathbf{L}_{1,n}(Q) + 2\nu \mathbf{L}_{2,n}(Q)}{\varsigma_Q \rho(1-3\rho)} \,. \tag{3.17}$$

This theorem is proved in [18].

Now we will obtain the oracle inequality for the model (1.1), (2.1). To write down the oracle inequality in this case, one needs the following parameters

$$\lambda_1 = \lambda \varrho_1^2 + (\lambda \varrho_2)^2 \quad \text{and} \quad \lambda_2 = \varrho_1^2 \varrho^* + \lambda \varrho_2^2 \,. \tag{3.18}$$

Moreover, we set

$$\mathbf{M}^* = 4\varrho_1^2 + \varrho_2^2 \mathbf{D}_1^* + 80\lambda_2 + 12\mathbf{D}_2^* + 21\varrho_3 \tag{3.19}$$

where $\mathbf{D}_1^* = 4\lambda\varrho_1^2 + 7\lambda^2\varrho_2^2$, $\mathbf{D}_2^* = 4\varrho_1^2\varrho^* + \varrho_2^2\mathbf{D}_1^* + 23\lambda_2$ and $\varrho_3 = \lambda\varrho_2^4 \mathbf{E} Y_1^4$.

Theorem 3.2. *Let $\mathcal{Q}_n$ be a family of the distributions of the process* (2.1) *with the parameters meeting the conditions* (2.4). *Then, for any $n \geq 1$ and $0 < \rho < 1/3$, the estimator* (3.15) *satisfies, for each $Q \in \mathcal{Q}_n$, the oracle inequality* (3.16) *with*

$$\Psi_Q(n,\rho) = \frac{6\varrho^*_{max}\nu + 4\varrho^*_{max}\mathbf{L}_1^* + 56\nu\mathbf{M}^*}{\varrho^*_{min}\rho(1-3\rho)} \,, \tag{3.20}$$

where

$$\mathbf{L}_1^* = 2(1 + a_{max}(a_{max}+1))\varrho^*_{max} \,.$$

Proof of this theorem is given in Section 5.

Remark 3.1. *Note that the term* (3.20) *does not depend on the parameters specifying the distribution family $\mathcal{Q}_n$. This means that the oracle inequality* (3.16) *is uniform over stability region for the process* (2.1) *including the stability bound, i.e. the case $a = 0$.*

To obtain the oracle inequality for the robust risks one has to impose additional conditions on the distribution family $\mathcal{Q}_n$ in (1.6). To this end, we introduce the family of distributions Q on $\mathcal{D}[0,n]$ with the growth restriction on $\mathbf{L}_{1,n}(Q) + \mathbf{L}_{2,n}(Q)$, i.e.

$$\mathcal{P}_n^* = \left\{ Q \in \mathcal{P}_n \,:\, \mathbf{L}_{1,n}(Q) + \mathbf{L}_{2,n}(Q) \leq l_n \right\} \,, \tag{3.21}$$

where $\mathcal{P}_n$ denotes the family of all probability measures on $\mathcal{D}[0,n]$, l_n is a slowly increasing positive function, i.e. $l_n \to +\infty$ as $n \to +\infty$ and for any $\delta > 0$

$$\lim_{n\to\infty} \frac{l_n}{n^\delta} = 0\,.$$

In the sequel we use the following condition
$\mathbf{H}_0$) *Assume the distribution family* $\mathcal{Q}_n$ *is a subset of the class* (3.21), *i.e.* $\mathcal{Q}_n \subseteq \mathcal{P}^*_n$, *such that*

$$\begin{aligned} 0 < \varsigma_* := \inf_{Q\in\mathcal{Q}_n} \varsigma_Q \le \sup_{Q\in\mathcal{Q}_n} \varsigma_Q := \varsigma^* \quad &< \infty\,; \\ \sigma^* := \inf_{Q\in\mathcal{Q}_n} \sigma_Q \quad &< \infty\,. \end{aligned} \tag{3.22}$$

Theorem 3.3. *Assume that the family* $\mathcal{Q}_n$ *in the robust risk* (1.6) *satisfies the condition* $\mathbf{H}_0$). *Then for any* $n \ge 1$ *and* $0 < \rho < 1/3$, *the estimator* (3.15) *satisfies the oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \le \frac{1+3\rho-2\rho^2}{1-3\rho} \min_{\gamma\in\Gamma} \mathcal{R}^*(\widehat{S}_\gamma, S) + \frac{1}{n}\,\mathcal{B}^*(n,\rho) \tag{3.23}$$

where

$$\mathcal{B}^*(n,\rho) = \Psi^*(n,\rho) + \frac{6\mu}{1-3\rho} \sup_{Q\in\mathcal{Q}_n} \mathbf{E}_{Q,S}|\widehat{\sigma}_n - \varsigma_Q|$$

and

$$\Psi^*(n,\rho) = \frac{2\varsigma^*\sigma^*\nu + 4\varsigma^* l_n + 2\nu l_n}{\varsigma_*\rho(1-3\rho)}\,.$$

3.1 Estimation of ς_Q

Now we consider the case of unknown quantity σ in the condition $\mathbf{C}_1$). One can estimate σ as

$$\widehat{\sigma}_n = \sum_{j=l}^{n} \widehat{\theta}^2_{j,n} \quad \text{with} \quad l = [\sqrt{n}] + 1\,. \tag{3.24}$$

Proposition 3.4. *Assume that the family distribution $\mathcal{Q}_n$ satisfies the condition $\mathbf{H}_0$) and the unknown function $S(\cdot)$ in the model* (1.1) *is continuously differentiable for $0 \le t < 1$ such that*

$$|\dot{S}|_1 = \int_0^1 |\dot{S}(t)|\mathrm{d}t < +\infty\,. \tag{3.25}$$

Then, for any $n \ge 1$,

$$\sup_{Q\in\mathcal{Q}_n} \mathbf{E}_{Q,S}|\widehat{\sigma}_n - \varsigma_Q| \le \frac{\kappa_n^*(S)}{\sqrt{n}} \tag{3.26}$$

where

$$\kappa_n^*(S) = 4|\dot{S}|_1^2 + \varsigma^* + \sqrt{l_n} + \frac{4|\dot{S}|_1\sqrt{\sigma^*}}{n^{1/4}} + \frac{l_n}{n^{1/2}}\,.$$

Proof. Substituting (3.4) in (3.24) yields

$$\widehat{\sigma}_n = \sum_{j=l}^n \theta_j^2 + \frac{2}{\sqrt{n}}\sum_{j=l}^n \theta_j\xi_{j,n} + \frac{1}{n}\sum_{j=l}^n \xi_{j,n}^2\,. \tag{3.27}$$

Further, denoting

$$x'_j = \mathbf{1}_{\{l\le j\le n\}} \quad \text{and} \quad x''_j = \frac{1}{\sqrt{n}}\mathbf{1}_{\{l\le j\le n\}}\,,$$

we represent the last term in (3.27) as

$$\frac{1}{n}\sum_{j=l}^n \xi_{j,n}^2 = \frac{1}{n}B_{1,n}(x') + \frac{1}{\sqrt{n}}\,B_{2,n}(x'') + \frac{n-l+1}{n}\varsigma_Q\,,$$

where

$$B_{1,n}(x) = \sum_{j=1}^\infty x_j\,\left(\mathbf{E}_{Q,S}\,\xi_{j,n}^2 - \varsigma_Q\right) \quad \text{and} \quad B_{2,n}(x) = \sum_{j=1}^\infty x_j\,(\xi_{j,n}^2 - \mathbf{E}_{Q,S}\xi_{j,n}^2)\,.$$

Combining these equations leads to the inequality

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \varsigma_Q| \le \sum_{j\ge l}\theta_j^2 + \frac{2}{\sqrt{n}}\mathbf{E}_{Q,S}|\sum_{j=l}^n \theta_j\xi_{j,n}|$$
$$+\frac{1}{n}|B_{1,n}(x')| + \frac{1}{\sqrt{n}}\,\mathbf{E}_{Q,S}\,|B_{2,n}(x'')| + \frac{l-1}{n}\varsigma^*\,.$$

By Lemma A.3 and the conditions $\mathbf{C}_1$), $\mathbf{C}_2$), one gets

$$\mathbf{E}_{Q,S}|\widehat{\sigma}_n - \varsigma_Q| \le \sum_{j\ge l} \theta_j^2 + \frac{2}{\sqrt{n}} \mathbf{E}_{Q,S}|\sum_{j=l}^{n} \theta_j \xi_{j,n}| + \frac{\mathbf{L}_{1,n}(Q)}{n} + \frac{\mathbf{L}_{2,n}(Q)}{\sqrt{n}} + \frac{\varsigma^*}{\sqrt{n}}\,.$$

In view of the inequality (1.3), the last term can be estimated as

$$\mathbf{E}_{Q,S}|\sum_{j=l}^{n} \theta_j \xi_{j,n}| \le \sqrt{\sigma_Q \sum_{j=l}^{n} \theta_j^2} \le \sqrt{\sigma^*}|\dot{S}|_1 \frac{2}{\sqrt{l}}\,.$$

By applying the inequalities (3.22) we obtain the upper bound (3.26). Hence Proposition 3.4.

□

Theorem 3.1 and Proposition 3.4 imply the following result.

Theorem 3.5. *Assume that the family distribution $\mathcal{Q}_n$ satisfies the condition $\mathbf{H}_0$) and the unknown function S is continuously differentiable satisfying the condition* (3.25). *Then, for any $n \ge 1$ and $0 < \rho < 1/3$, the model selection procedure* (3.15) *with the estimator* (3.24) *satisfies the oracle inequality*

$$\mathcal{R}^*(\widehat{S}_*, S) \le \frac{1 + 3\rho - 2\rho^2}{1 - 3\rho} \min_{\gamma\in\Gamma} \mathcal{R}^*(\widehat{S}_\gamma, S) + \frac{1}{n}\, \mathcal{B}_1^*(n,\rho)\,, \tag{3.28}$$

where

$$\mathcal{B}_1^*(n,\rho) = \Psi^*(n,\rho) + \frac{6\mu\kappa_n^*(S)}{(1-3\rho)\sqrt{n}}\,.$$

3.2 Specification of weights in the model selection procedure (3.15)

We will specify the weight coefficients $(\gamma(j))_{j\ge 1}$ in the way proposed in [11] for a heteroscedastic discrete time regression model. Consider a numerical grid of the form

$$\mathcal{A}_n = \{1, \ldots, k^*\} \times \{t_1, \ldots, t_m\}\,, \tag{3.29}$$

where $t_i = i\varepsilon$ and $m = [1/\varepsilon^2]$. We assume that both parameters $k^* \ge 1$ and $0 < \varepsilon \le 1$ are functions of n, i.e. $k^* = k^*(n)$ and $\varepsilon = \varepsilon(n)$,

such that

$$
\begin{cases}
\lim_{n\to\infty} k^*(n) = +\infty\,, \quad \lim_{n\to\infty} \dfrac{k^*(n)}{\ln n} = 0\,, \\[2ex]
\lim_{n\to\infty} \varepsilon(n) = 0 \quad \text{and} \quad \lim_{n\to\infty} n^{\delta}\varepsilon(n) \ = +\infty
\end{cases}
\tag{3.30}
$$

for any $\delta > 0$. One can take, for example,

$$
\varepsilon(n) = \frac{1}{\ln(n+1)} \quad \text{and} \quad k^*(n) = \sqrt{\ln(n+1)}\,.
$$

For each $\alpha = (\beta, t) \in \mathcal{A}_n$, we introduce the weight sequence $\gamma_\alpha = (\gamma_\alpha(j))_{j\ge 1}$ given as

$$
\gamma_\alpha(j) = \mathbf{1}_{\{1\le j\le j_0\}} + \left(1 - (j/\omega_\alpha)^{\beta}\right) \mathbf{1}_{\{j_0 < j\le \omega_\alpha\}}
\tag{3.31}
$$

where $j_0 = j_0(\alpha) = [\omega_\alpha/\ln n]$,

$$
\omega_\alpha = (\tau_\beta\, t\, n)^{1/(2\beta+1)} \quad \text{and} \quad \tau_\beta = \frac{(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}\,.
$$

We set

$$
\Gamma \ = \ \{\gamma_\alpha\,,\, \alpha \in \mathcal{A}_n\}\,.
\tag{3.32}
$$

It will be noted that in this case $\nu = k^* m$.

Remark 3.2. *It will be observed that the specific form of weights* (3.31) *was proposed by Pinsker [24] for the filtration problem with known smoothness of regression function observed with an additive gaussian white noise in the continuous time. Nussbaum [23] used these weights for the gaussian regression estimation problem in discrete time.*

The minimal mean square risk, called the Pinsker constant, is provided by the weight least squares estimate with the weights where the index α depends on the smoothness order of the function S. In this case the smoothness order is unknown and, instead of one estimate, one has to use a whole family of estimates containing in particular the optimal one.

The problem is to study the properties of the whole class of estimates. Below we derive an oracle inequality for this class which yields the best mean square risk up to a multiplicative and additive constants provided that the the smoothness of the unknown function S is not available. Moreover, it will be shown that the multiplicative constant tends to unity and the additive one vanishes as $n \to \infty$ with the rate higher than any minimax rate.

In view of the assumptions (3.30), for any $\delta > 0$, one has

$$\lim_{n\to\infty} \frac{\nu}{n^{\delta}} = 0\,.$$

Moreover, by (3.31) for any $\alpha \in \mathcal{K}_n$

$$\sum_{j=1}^{\infty} \mathbf{1}_{\{\gamma_\alpha(j)>0\}} \le \omega_\alpha\,.$$

Therefore, taking into account that $A_\beta \le A_1 < 1$ for $\beta \ge 1$, we get

$$\mu = \mu_n \le (n/\varepsilon)^{1/3}\,.$$

Therefore, for any $\delta > 0$,

$$\lim_{n\to\infty} \frac{\mu_n}{n^{1/3+\delta}} = 0\,.$$

To study the asymptotic behaviour of the term $\mathcal{B}_1^*(n,\rho)$ we assume that the parameter ρ in the cost function (3.11) depends on n, i.e. $\rho = \rho_n$ such that $\rho_n \to 0$ as $n \to \infty$ and for any $\delta > 0$

$$\lim_{n\to\infty} n^{\delta} \rho_n = 0\,. \tag{3.33}$$

Applying this limiting relation to the analysis of the asymptotic behavior of the additive term $\mathcal{D}_n(\rho)$ in (3.28) one comes to the following result.

Theorem 3.6. *Assume that the family distribution $\mathcal{Q}_n$ satisfies the condition $\mathbf{H}_0$) and the unknown function S is continuously differentiable satisfying the condition* (3.25)*. Then, for any $n \ge 1$, the model selection procedure* (3.15), (3.33), (3.24), (3.32) *satisfies the oracle inequality* (3.28) *with the additive term $\mathcal{B}_1^*(n,\rho)$ obeying, for any $\delta > 0$, the following limiting relation*

$$\lim_{n\to\infty} \frac{\mathcal{B}_1^*(n,\rho_n)}{n^{\delta}} = 0\,.$$

4 Stochastic integrals with respect to the process (2.1)

In this Section we establish some properties of a stochastic integral

$$I_t(f) = \int_0^t f_s \mathrm{d}\xi_s \quad 0 \le t \le n\,, \tag{4.1}$$

with respect to to the process (2.1). We will need some notations. Let us denote

$$\varepsilon_f(t) = a \int_0^t e^{a(t-v)}\, f(v)\,(1 + e^{2av})\,\mathrm{d}v\,, \tag{4.2}$$

where f is $[0,+\infty) \to \mathbb{R}$ function integrated on any finite interval. We introduce also the following transformation

$$\tau_{f,g}(t) = \frac{1}{2} \int_0^t \Big(2f(s)g(s) + \varepsilon^*_{f,g}(s)\Big)\,\mathrm{d}s \tag{4.3}$$

of square integrable $[0,+\infty) \to \mathbb{R}$ functions f and g. Here

$$\varepsilon^*_{f,g}(t) = f(t)\varepsilon_g(t) + \varepsilon_f(t)g(t)\,.$$

It will be noted that

$$a\tau_{f,1}(t) = \frac{1}{2}\varepsilon_f(t) \quad \text{and} \quad a\tau_{1,1}(t) = \frac{1}{2}\left(e^{2at} - 1\right)\,. \tag{4.4}$$

Proposition 4.1. *If f and g are from $\mathcal{L}_2[0,n]$ then*

$$\mathbf{E}\, I_t(f) I_t(g) = \varrho^*\, \tau_{f,g}(t) \tag{4.5}$$

where ϱ^* is given in (2.4).

Proof. Noting that the process $I_t(f)$ satisfies the stochastic equation

$$\mathrm{d}I_t(f) = af(t)\xi_t \mathrm{d}t + f(t)\mathrm{d}u_t\,, \quad I_0(f) = 0\,,$$

and applying the Ito formula one obtains (4.5). Hence Proposition 4.1. $\square$

Further, for integrated $[0,+\infty) \to \mathbb{R}$ functions f and g, we define the $[0,+\infty) \times [0,+\infty) \to \mathbb{R}$ function

$$D_{f,g}(x,z) = \int_0^x L^*_{f,g}(y,z)\,\mathrm{d}y + f(z)g(z)\,, \tag{4.6}$$

where $L^*_{f,g}(y,z) = g(y+z)L_f(y,z) + f(y+z)L_g(y,z)$;

$$L_f(x,z) = ae^{ax}\left(f(z) + a\int_0^x e^{av} f(v+z)\,\mathrm{d}v\right).$$

Proposition 4.2. *Let $\mathcal{G}_k = \sigma\{T_1, \ldots, T_k\}$, where $k \geq 1$, be σ-algebra generated by the stopping times* (2.3)*, f and g be bounded left-continuous $[0,\infty) \times \Omega \to \mathbb{R}$ functions measurable with respect to $\mathcal{B}[0,+\infty) \bigotimes \mathcal{G}_k$ (the product σ algebra created by $\mathcal{B}[0,+\infty)$ and $\mathcal{G}_k$). Then*

$$\mathbf{E}\left(I_{T_k-}(f)|\mathcal{G}_k\right) = 0$$

and

$$\mathbf{E}\left(I_{T_k-}(f)\, I_{T_k-}(g)|\mathcal{G}_k\right) = \varrho_1^2 \tau_{f,g}(T_k) + \varrho_2^2 \sum_{l=1}^{k-1} D_{f,g}(T_k - T_l, T_l).$$

Proof. By the Ito formula one has

$$\begin{aligned} I_t(f)\, I_t(g) = & \int_0^t (\varrho_1^2 f(s)g(s) + a(f(s)I_s(g) + g(s)I_s(f))\xi_s)\mathrm{d}s \\ & + \varrho_2^2 \sum_{l\geq 1} f(T_l)\, g(T_l)\, Y_l^2 \mathbf{1}_{\{T_l \leq t\}} \\ & + \int_0^t (f(s)I_{s-}(g) + g(s)I_{s-}(f)))\mathrm{d}u_s\,. \end{aligned} \tag{4.7}$$

Taking the conditional expectation $\mathbf{E}\,(\cdot|\mathcal{G}_k)$, on the set $\{T_k > t\}$, yields

$$\begin{aligned} \mathbf{E}\,(I_t(f)\, I_t(g)|\mathcal{G}_k) = & \int_0^t \varrho_1^2 f(s)g(s)\mathrm{d}s + \varrho_2^2 \sum_{l\geq 1} f(T_l)\, g(T_l)\; \mathbf{1}_{\{T_l \leq t\}} \\ & + a\int_0^t \left(f(s)\mathbf{E}(I_s(g)\xi_s|\mathcal{G}_k) + g(s)\mathbf{E}(I_s(f)\xi_s|\mathcal{G}_k)\right)\mathrm{d}s\,. \end{aligned}$$

Now to calculate the function $Z_t = \mathbf{E}(I_t(f)\xi_t|\mathcal{G}_k)$ we put $g = 1$. Taking into account that

$$\mathbf{E}(\xi_t^2|\mathcal{G}_k) = \frac{\varrho_1^2}{2a}(e^{2at} - 1) + \varrho_2^2 \sum_{l\geq 1} e^{2a(t-T_l)}\, \mathbf{1}_{\{T_l \leq t\}}\,,$$

one obtains, for $T_{j-1} \le t < T_j$,

$$\dot{Z}_t = aZ_t + f(t)\psi_t \,,$$

where

$$\psi_t = \frac{\varrho_1^2}{2}\left(1+e^{2at}\right) + \varrho_2^2 a \sum_{l\ge 1} e^{2a(t-T_l)}\mathbf{1}_{\{T_l\le t\}} \,.$$

Therefore, for $T_{j-1} \le t < T_j$,

$$Z_t = e^{a(t-T_{j-1})} Z_{T_{j-1}} + \int_{T_{j-1}}^{t} e^{a(t-s)} f(s)\psi_s \mathrm{d}s \,.$$

¿From here and (4.6) with $g=1$ one has

$$Z_{T_j} = e^{a(T_j - T_{j-1})} Z_{T_{j-1}} + \eta_j \,, \quad Z_{T_0} = Z_0 = 0 \,, \tag{4.8}$$

where

$$\eta_j = \int_{T_{j-1}}^{T_j} e^{a(T_j - s)} f(s)\psi_s \mathrm{d}s + \varrho_2^2 f(T_j) \,.$$

Solving the equation (4.7) one obtains

$$Z_{T_j} = \int_0^{T_j} e^{a(T_j - s)} f(s)\psi_s \mathrm{d}s + \varrho_2^2 \sum_{l=1}^{j} e^{a(T_j - T_l)} f(T_l) \,.$$

Therefore,

$$Z_t = \int_0^t e^{a(t-s)} f(s)\psi_s \mathrm{d}s + \varrho_2^2 \sum_{l\ge 1} e^{a(t-T_l)} f(T_l)\mathbf{1}_{\{T_l\le t\}} \,,$$

i.e

$$a\mathbf{E}(I_t(f)\xi_t|\mathcal{G}_k) = \frac{\varrho_1^2}{2}\varepsilon_f(t) + \varrho_2^2 \sum_{j\ge 1} L_f(t-T_j, T_j)\,\mathbf{1}_{\{T_j\le t\}}.$$

¿From here one comes to the desired equality. Hence Proposition 4.2. $\square$

Proposition 4.3. *Let F, f and g be non random bounded left-continuous $[0,\infty) \to \mathbb{R}$ functions. Then*

$$\mathbf{E} \sum_{k\ge 1} F(T_k)\, I_{T_{k-}}(f)\, I_{T_{k-}}(g)\, \mathbf{1}_{\{T_k\le t\}} = \int_0^t F(v)\, H_{f,g}(v)\, \mathrm{d}v \,,$$

where

$$H_{f,g}(t) = \lambda\varrho_1^2\,\tau_{f,g}(t) + (\lambda\varrho_2)^2\int_0^t D_{f,g}(t-z,z)\mathrm{d}z\,.$$

Proof. We have

$$\iota(t) = \mathbf{E}\sum_{k\ge1} F(T_k)\,I_{T_{k-}}(f)\,I_{T_{k-}}(g)\,\mathbf{1}_{\{T_k\le t\}}\,.$$

By Proposition 4.2 one gets

$$\begin{aligned}
\iota(t) &= \varrho_1^2\mathbf{E}\sum_{k\ge1} F(T_k)\,\tau_{f,g}(T_k)\,\mathbf{1}_{\{T_k\le t\}}\\
&\quad+\varrho_2^2\,\mathbf{E}\sum_{k\ge1} F(T_k)\,\sum_{l=1}^{k-1} D_{f,g}(T_k-T_l,T_l)\mathbf{1}_{\{T_k\le t\}}\\
&:= \iota_1(t)+\iota_2(t)\,,
\end{aligned}$$

where

$$\iota_1(t) = \lambda\int_0^t\sum_{l\ge1} F(z)\tau_{f,g}(z)\frac{(\lambda z)^{l-1}}{(l-1)!}e^{-\lambda z}\mathrm{d}z = \int_0^t F(z)\tau_{f,g}(z)\mathrm{d}z\,.$$

To calculate $\iota_2(t)$ we note that

$$\iota_2(t) = \mathbf{E}\sum_{l\ge1}\mathbf{1}_{\{T_l\le t\}}\sum_{k\ge l+1} F(T_k)D_{f,g}(T_k-T_l,T_l)\mathbf{1}_{\{T_k\le t\}}\,.$$

Taking into account that T_k-T_l is independent of T_l for any $k>l$ we obtain

$$\begin{aligned}
\iota_2(t) &= \lambda\mathbf{E}\sum_{l\ge1}\mathbf{1}_{\{T_l\le t\}}\int_0^{t-T_l}\sum_{k\ge l+1} F(z+T_l)D_{f,g}(z,T_l)\,\frac{(\lambda z)^{k-l-1}}{(k-l-1)!}e^{-\lambda z}\mathrm{d}z\\
&= \lambda\mathbf{E}\sum_{l\ge1}\mathbf{1}_{\{T_l\le t\}}\int_0^{t-T_l} F(z+T_l)D_{f,g}(z,T_l)\,\mathrm{d}z\\
&= \lambda^2\int_0^t\int_0^{t-x}\big(F(z+x)\,D_{f,g}(z,x)\mathrm{d}z\big)\,\mathrm{d}x\,.
\end{aligned}$$

□

Note that

$$aH_{f,1}(t) = \frac{\lambda_1}{2}\,\varepsilon_f(t) \quad \text{and} \quad aH_{1,1}(t) = \frac{\lambda_1}{2}\left(e^{2at} - 1\right)\,, \tag{4.9}$$

where λ_1 given in (3.18). Now we set

$$\widetilde{I}_t(f) = I_t^2(f) - \mathbf{E}\,I_t^2(f)\,. \tag{4.10}$$

Further we need the following correlation measures for two integrated $[0,+\infty) \to \mathbb{R}$ functions f and g

$$\varpi_{f,g} = \max_{0\le v\le n}\,\max_{0\le t\le n-v}\,|\int_0^t f(u+v)g(u)\mathrm{d}u| \tag{4.11}$$

and

$$\varpi^*_{f,g} = \max\left(\varpi_{f,g}, \varpi_{g,f}\right)\,. \tag{4.12}$$

For any bounded $[0,\infty) \to \mathbb{R}$ function f we introduce the following uniform norm

$$\|f\|_* = \sup_{0\le t\le n}\,|f(t)|\,.$$

To check the condition $\mathbf{C}_2$) one needs the following non-asymptotic upper bound

Theorem 4.4. *For any bounded left-continuous $[0,\infty) \to \mathbb{R}$ functions f, g*

$$|\mathbf{E}\widetilde{I}_n(f)\widetilde{I}_n(g)| \le n\mathbf{M}^*\left(\varpi^*_{f,g} + \|f\|_*\|g\|_*\right)\|f\|_*\|g\|_*\,, \tag{4.13}$$

where $\mathbf{M}^*$ is defined in (3.19).

Proof. By the Ito formula one comes to the following stochastic equation

$$\mathrm{d}\widetilde{I}_t(f) = 2av_t(f)f(t)\mathrm{d}t + \mathrm{d}M_t(f)\,, \quad \widetilde{I}_0(f) = 0\,,$$

with $v_t(f) = I_t(f)\xi_t - \mathbf{E}\,I_t(f)\xi_t$,

$$\mathrm{d}M_t(f) = 2I_{t-}(f)f(t)\mathrm{d}u_t + \varrho_2^2 f^2(t)\mathrm{d}m_t\,, \quad M_0(f) = 0\,.$$

In view of Propositions 4.1–4.3, one finds

$$\mathbf{E}\,[\widetilde{I}(f), \widetilde{I}(g)]_t = \mathbf{E}\,[M(f), M(g)]_t = \int_0^t V_{f,g}(s)\,\mathrm{d}s\,, \tag{4.14}$$

where $V_{f,g}(t) = 4f(t)g(t)G_{f,g}(t) + \varrho_3 f^2(t)\, g^2(t)$ and $G_{f,g} = \varrho_1^2 \varrho^* \tau_{f,g}(t) + \varrho_2^2 H_{f,g}(t)$. Note that Lemma A.1 implies

$$\max_{0\le t\le n} |G_{f,g}(t)| \,\le\, (4\varrho_1^2\varrho^* + \varrho_2^2 \mathbf{D}_1^*)\varpi_{f,g}^* \,. \tag{4.15}$$

One can easily check that

$$V_{1,1}(t) = 2\lambda_2 \frac{e^{2at}-1}{a} + \varrho_3 \,. \tag{4.16}$$

The constants λ_2, ϱ_3 and $\mathbf{D}_1^*$ are given in (3.18)-(3.19). Moreover, by the Ito formula we get

$$\mathrm{d}v_t(f) = av_t(f)\mathrm{d}t + af(t)\zeta_t \mathrm{d}t + \mathrm{d}K_t(f)\,, \quad v_0(f) = 0\,,$$

where $\zeta_t = \xi_t^2 - \mathbf{E}\,\xi_t^2$,

$$K_t(f) = \int_0^t I_{s-}^*(f)\,\mathrm{d}u_s + \int_0^t \varrho_2^2 f(s)\mathrm{d}m_s \tag{4.17}$$

with

$$I_t^*(f) = I_t(f) + f(t)\xi_t \quad \text{and} \quad m_t = \sum_{0\le s\le t} \Delta z_s^2 - \lambda t\,.$$

Proposition 4.1 implies

$$\mathbf{E}\, I_t^*(f) I_t^*(g) = \varrho^*\, \tau_{f,g}^*(t)\,,$$

where

$$\tau_{f,g}^*(t) = \tau_{f,g}(t) + f(t)\tau_{1,g}(t) + g(t)\tau_{f,1}(t) + f(t)g(t)\,\tau_{1,1}(t)\,.$$

¿From (4.3)–(4.4) it follows that

$$\tau_{f,g}^*(t) = \tau_{f,g}(t) + \frac{\varepsilon_{f,g}^*(t) + f(t)g(t)(e^{2at}-1)}{2a}\,. \tag{4.18}$$

By applying Proposition 4.3 one finds

$$\mathbf{E}\sum_{k\ge1} I_{T_k-}^*(f)\, I_{T_k-}^*(g)\, \mathbf{1}_{\{T_k\le t\}} = \int_0^t H_{f,g}^*(v)\,\mathrm{d}v\,,$$

where

$$H_{f,g}^*(t) = H_{f,g}(t) + f(t)H_{1,g}(t) + g(t)H_{f,1}(t) + f(t)g(t)\, H_{1,1}(t)\,.$$

¿From here and (4.9) we get

$$H^*_{f,g}(t) = H_{f,g}(t) + \lambda_1 \frac{\varepsilon^*_{f,g}(t) + f(t)g(t)(e^{2at} - 1)}{2a} \,. \tag{4.19}$$

Taking this into account, we calculate that, for any square integrated functions f and g,

$$\mathbf{E}\,[K(f)\,,\,K(g)]_t = \int_0^t \Big(G^*_{f,g}(s) + \varrho_3\, f(s)\, g(s)\Big)\, \mathrm{d}s\,, \tag{4.20}$$

where

$$G^*_{f,g}(t) = \varrho_1^2 \varrho^* \tau^*_{f,g}(t) + \varrho_2^2\, H^*_{f,g}(t)\,.$$

Further by applying Propositions 4.1–4.3 we obtain

$$\mathbf{E}\,[K(f), M(g)]_t = \int_0^t U_{f,g}(s)\mathrm{d}s\,, \tag{4.21}$$

where

$$U_{f,g}(s) = 2g(s)G_{f,g}(s) + 2g(s)f(s)G_{1,g}(s) + \varrho_3 f(s)g^2(s)\,.$$

By the Ito formula one finds for $t \geq 0$

$$\mathbf{E}\,\widetilde{I}_t(f)\,\widetilde{I}_t(g) = \mathbf{E}\,[\widetilde{I}(f)\,,\,\widetilde{I}(g)]_t + 2\int_0^t \big(f(s)\,\mathcal{T}_{f,g}(s) + g(s)\,\mathcal{T}_{g,f}(s)\big)\,\mathrm{d}s\,, \tag{4.22}$$

where $\mathcal{T}_{f,g}(t) = a\mathbf{E}\, v_t(f)\,\widetilde{I}_t(g)$. Since for $g = 1$ the processes $\widetilde{I}_t(g)$ and $v_t(g)$ coincide with ζ_t, i.e. $\widetilde{I}_t(1) = v_t(1) = \zeta_t$, (4.17) implies

$$\mathbf{E}\zeta_t^2 = \int_0^t e^{4a(t-s)}\, V_{1,1}(s)\mathrm{d}s = e^{4at}\frac{2\lambda_2 + a\varrho_3}{4a^2} + e^{2at}\frac{\lambda_2}{a^2} + \frac{2\lambda_2 - a\varrho_3}{4a^2}\,. \tag{4.23}$$

We define the function

$$A_f(t) = \int_0^t e^{3a(t-s)}\,\big(f(s)a^2\,\mathbf{E}\,\zeta_s^2 + \kappa_f(s)\big)\,\mathrm{d}s\,, \tag{4.24}$$

where

$$\kappa_f(t) = \lambda_2\left(\varepsilon_f(t) + f(t)\left(e^{2at} - 1\right)\right) + a\varrho_3 f(t)\,.$$

Denote

$$V_{f,g}(s) = A^*_{f,g}(s) + G^*_{f,g}(s) + \varrho_3\, f(s)g(s)\,, \tag{4.25}$$

where $A^*_{f,g}(t) = g(s)A_f(s) + f(s)A_g(s)$.
To calculate the function $\mathcal{T}_{f,g}(t)$, we note that, by the Ito formula and (4.16),

$$\begin{aligned}\mathrm{d}\mathbf{E}\, v_t(f)\widetilde{I}_t(g) &= a(\mathbf{E}v_t(f)\widetilde{I}_t(g))\mathrm{d}t + 2ag(t)(\mathbf{E}v_t(f)\, v_t(g))\mathrm{d}t \\ &\quad + af(t)(\mathbf{E}\zeta_t\widetilde{I}_t(g))\mathrm{d}t + U_{f,g}(t)\mathrm{d}t\,. \qquad (4.26)\end{aligned}$$

Substituting here $f = 1$, and then taking into account (A.2) yield

$$\mathbf{E}\zeta_t\widetilde{I}_t(g) = \int_0^t e^{2a(t-s)}\left(A^*_{g,g}(s) + U_{1,g}(s)\right)\mathrm{d}s\,.$$

Furthermore, by (A.2)

$$\begin{aligned}\mathbf{E}\, v_t(f)\widetilde{I}_t(g) &= 2a\int_0^t e^{a(t-s)}\, g(s)\ (\mathbf{E}v_s(f)v_s(g))\ \mathrm{d}s \\ &\quad + a\int_0^t e^{a(t-s)} f(s)\ \left(\mathbf{E}\zeta_s\,\widetilde{I}_s(g)\right)\ \mathrm{d}s + \int_0^t e^{a(t-s)}\, U_{f,g}(s)\mathrm{d}s\,.\end{aligned}$$

Therefore, for any bounded left-continuous $[0,+\infty) \to \mathbb{R}$ functions f and g, one finds

$$\mathcal{T}_{f,g}(t) = a\int_0^t e^{a(t-s)}\left(g(s)\mathcal{V}_{f,g}(s) + f(s)\mathcal{K}_g(s) + U_{f,g}(s)\right)\mathrm{d}s\,, \quad (4.27)$$

where

$$\mathcal{V}_{f,g}(t) = 2a\int_0^t e^{2a(t-s)}\, V_{f,g}(s)\mathrm{d}s$$

and

$$\mathcal{K}_g(t) = a\int_0^t e^{2a(t-s)}\left(A^*_{g,g}(s) + U_{1,g}(s)\right)\mathrm{d}s\,.$$

¿From (4.17) and (4.27), it follows that

$$|\mathbf{E}\,\widetilde{I}_n(f)\,\widetilde{I}_n(g)| \le n\|V_{f,g}\|_* + 2n\|f\|_*\|\mathcal{T}_{f,g}\|_* + 2n\|g\|_*\|\mathcal{T}_{g,f}\|_*\,.$$

Now by applying the inequality (A.10) one gets

$$\|V_{f,g}\|_* \le \left((4\varrho_1^2 + \varrho_2^2\mathbf{D}_1^*)\varpi^*_{f,g} + \varrho_3\|f\|_*\|g\|_*\right)\|f\|_*\|g\|_*\,.$$

Note that Lemmas A.6–A.8 imply

$$\|\mathcal{T}_{f,g}\|_* \le \left((20\lambda_2 + 3\mathbf{D}_2^*)\varpi^*_{f,g} + 5\varrho_3\|f\|_*\|g\|_*\right)\|g\|_*\,.$$

¿From here one comes to the upper bound (4.12).
□

5 Proof of Theorem 3.2

First we note that Proposition 4.1 implies the inequality (1.3) with $\sigma_Q = 3\varrho^*$. Therefore due to the conditions (2.4) one obtains $\sigma^* = 3\varrho_{max}$. Now we verify Conditions $\mathbf{C}_1$) and $\mathbf{C}_2$) for the family of processes (2.1) satisfying the conditions (2.4). To begin with we note that

$$\mathbf{E}_{Q,S}\xi^2_{j,n} = \varrho^* \left(1 + \frac{a}{n}\int_0^n e^{av}\,\Upsilon_j(v)\mathrm{d}v\right)\,,$$

where

$$\Upsilon_j(v) = \int_v^n \phi_j(t)\,\phi_j(t-v)\,\left(1+e^{2a(t-v)}\right)\mathrm{d}t\,.$$

If $j=1$, one has

$$|\mathbf{E}_{Q,S}\xi^2_{1,n} - \varrho^*| \le 2\varrho^*\,. \tag{5.1}$$

Since for the trigonometric basis (3.1) for $j \ge 2$

$$\phi_j(t)\,\phi_j(t-v) = \cos(\gamma_j v) + (-)^j\cos(\gamma_j(2t-v))$$

where $\gamma_j = 2\pi[j/2]$, therefore,

$$\Upsilon_j(v) = \cos(\gamma_j v)F(v) + (-1)^j\,\Upsilon_{0,j}(v)\,,\quad F(v) = \int_0^{n-v}\left(1+e^{2at}\right)\mathrm{d}t$$

and

$$\Upsilon_{0,j}(v) = \int_0^{n-v}\cos(\gamma_j(2t+v))\left(1+e^{2at}\right)\mathrm{d}t\,.$$

Integrating by parts yields

$$\Upsilon_{0,j}(v) = \frac{e^{2a(n-v)}}{2\gamma_j}\,\sin(v\gamma_j) + \frac{a}{2\gamma_j^2}\,\Upsilon_{1,j}(v)$$

where

$$\Upsilon_{1,j}(v) = \cos(v\gamma_j)(e^{a(n-v)}-1) - a\int_0^{n-v}e^{at}\cos((2t+v)\gamma_j)\,\mathrm{d}t\,.$$

It is obvious, that $|\Upsilon_{1,j}(v)| \le 2$. Further we obtain

$$a\int_0^n e^{av}\,\Upsilon_j(v)\mathrm{d}v = a\int_0^n e^{av}\,F(v)\,\cos(v\gamma_j)\mathrm{d}v + a(-1)^j\int_0^n e^{av}\,\Upsilon_{0,j}(v)\,\mathrm{d}v$$

$$:= aD_1(n) + a(-1)^j D_2(n)\,.$$

Integrating by parts two times we find

$$D_1(n) = \frac{1}{\gamma_j^2}\left(e^{an}\dot{F}(n) - \dot{F}(0) - aF(0) - \int_0^n e^{av}F_1(v)\mathrm{d}v\right)\,,$$

where

$$F_1(v) = a^2F(v) + 2a\dot{F}(v) + \ddot{F}(v)\,.$$

This implies

$$|D_1(n)| \le \frac{1}{\gamma_j^2}\left(3n|a| + 10\right)\,.$$

Similarly, one gets

$$|D_2(n)| \le \frac{2}{\gamma_j^2}\,.$$

Thus, for $j \ge 2$,

$$|\mathbf{E}_{Q,S}\xi_{j,n}^2 - \varrho^*| \le \frac{15\,|a|(1+|a|)\varrho^*}{\pi^2 j^2}\,. \tag{5.2}$$

Therefore

$$\mathbf{L}_{1,n}(Q) \le 2(1 + |a|(|a|+1))\varrho^*$$

and taking into account the conditions (2.4) we get

$$\mathbf{L}_{1,n}(Q) \le \mathbf{L}_1^*\,, \tag{5.3}$$

where $\mathbf{L}_1^*$ is defined in (3.18). It means that the condition $\mathbf{C}_1$) holds with $\varsigma_Q = \varrho^*$. Moreover, by applying the conditions (2.4) we have $\varsigma^* = \varrho^*_{max}$ and $\varsigma_* = \varrho^*_{min}$.
To check the condition $\mathbf{C}_2$) we note that

$$\mathbf{E}_{Q,S}\left(\sum_{j=1}^{\infty} x_j\,(\xi_{j,n}^2 - \mathbf{E}_{Q,S}\xi_{j,n}^2)\right)^2 = \frac{1}{n^2}\sum_{i,j\ge1} x_i x_j\,\mathbf{E}_{Q,S}\widetilde{I}_n(\phi_i)\widetilde{I}_n(\phi_j)\,.$$

Therefore, in view of Theorem 4.4

$$\mathbf{E}_{Q,S}\left(\sum_{j=1}^{\infty} x_j\,(\xi_{j,n}^2 - \mathbf{E}_{Q,S}\xi_{j,n}^2)\right)^2 \le \frac{2\mathbf{M}^*}{n}\sum_{i,j\ge1}|x_i||x_j|(\varpi^*_{i,j} + 2)\,, \tag{5.4}$$

where $\varpi^*_{i,j} = \varpi^*_{\phi_i,\phi_j}$. To estimate this term we note, that for any $j \ge 1$,

$$\phi_j(v+u) = a_{j-1}(v)\phi_{j-1}(u) + a_j(v)\phi_j(u) + a_{j+1}(v)\phi_{j+1}(u)$$

$a_j(\cdot)$ are bounded functions with $|a_j(v)| \le 1$. Thus,

$$\varpi^*_{i,j} \le 3n\mathbf{1}_{\{|i-j|\le 1\}} + 3\mathbf{1}_{\{|i-j|\ge 2\}}\,.$$

Since $\sum_{j\ge1} x_j^2 \le 1$, therefore, the upper bound in (5.4) can be estimated as

$$\sum_{i,j\ge1} |x_i||x_j|(\varpi^*_{i,j}+2) \le 14n\,.$$

¿From here, it follows that

$$\mathbf{L}_{2,n}(Q) \le 28\mathbf{M}^*\,. \tag{5.5}$$

Hence Theorem 3.2.

□

6 Robust asymptotic efficiency

In this Section we show that the model selection procedure (3.15), (3.33), (3.24), (3.32) for estimating S in the model (1.1) is asymptotically efficient with respect to the robust risk (1.6). We assume that the unknown function S in the model (1.1) belongs to the Sobolev ball

$$W^k_r = \{f \in \mathcal{C}^k_{per}[0,1]\,,\, \sum_{j=0}^k \|f^{(j)}\|^2 \le r\}\,, \tag{6.1}$$

where $r>0\,,\ k\ge1$ are some parameters, $\mathcal{C}^k_{per}[0,1]$ is the set of k times continuously differentiable functions $f\,:\,[0,1]\to\mathbb{R}$ such that $f^{(i)}(0)=f^{(i)}(1)$ for all $0\le i\le k$. The functional class W^k_r can be written as an ellipsoid in l_2, i.e.

$$W^k_r = \{f \in \mathcal{C}^k_{per}[0,1]\,:\,\sum_{j=1}^\infty a_j\,\theta^2_j\,\le r\} \tag{6.2}$$

where $a_j = \sum_{i=0}^k (2\pi[j/2])^{2i}$.

We denote by Q_0 the distribution of Winer process with the scale parameter ς^* defined in (3.22).

$\mathbf{H}_1$) *Assume the distribution Q_0 belongs to the family $\mathcal{Q}_n$.*

In this Section we will show that the Pinsker constant for the robust risk (1.6) is given by the equation

$$R_k^* = ((2k+1)r)^{1/(2k+1)} \left(\frac{\varsigma^* k}{(k+1)\pi} \right)^{2k/(2k+1)} . \tag{6.3}$$

It is well known that the optimal (minimax) rate for the Sobolev ball W_r^k is $n^{2k/(2k+1)}$ (see, for example, [24], [23]).

We will see that asymptotically the robust risk (1.6) normalized by this rate is bounded from below by R_k^*, i.e. this bound can not be diminished if one considers the class of all admissible estimates for S. Let Π_n be the set of all estimators $\widehat{S}_n$ measurable with respect to the sigma-algebra $\sigma\{y_t\,,\,0 \le t \le n\}$ generated by the process (1.1).

Theorem 6.1. *Under the condition* $\mathbf{H}_1$)

$$\liminf_{n\to\infty} n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}_n^*(\widehat{S}_n, S) \ge R_k^* \,. \tag{6.4}$$

Proof of this theorem follows directly from Theorem 3.2 in [19].
Now we show that, under some conditions, the normalized robust risk for the model selection procedure is bounded from above by the same constant R_k^*.

Theorem 6.2. *Assume that, in model* (1.1)*, for each* $n \ge 1$ *the distribution of* $(\xi_t)_{0\le t\le n}$ *belongs to the family* $\mathcal{Q}_n$ *satisfying the conditions* $\mathbf{H}_0$)*. Then the robust risk* (1.6) *of the model selection procedure* $\widehat{S}_*$ *defined in* (3.33), (3.24), (3.32) *has the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n^*(\widehat{S}_*, S) \le R_k^* \,. \tag{6.5}$$

Theorem 6.1 and Theorem 6.2 imply the following result

Corollary 6.3. *Under the conditions* $\mathbf{H}_0$) *and* $\mathbf{H}_1$)

$$\lim_{n\to\infty} n^{2k/(2k+1)} \inf_{\widehat{S}_n\in\Pi_n} \sup_{S\in W_r^k} \mathcal{R}_n^*(\widehat{S}_n, S) = R_k^* \,. \tag{6.6}$$

Remark 6.1. *The equation* (6.6) *means that the parameter* R_k^* *defined by* (6.3) *is the Pinsker constant (see, for example, [24], [23]) for the model* (1.1)*. Moreover, the equality* (6.6) *means that the model selection procedure* (3.33), (3.24), (3.32) *is asymptotically robust efficient.*

7 Upper bound

7.1 Known smoothness

First we suppose that the parameters $k \geq 1$, $r > 0$ in (6.1) and ς^* in (3.22) are known. Let the family of admissible weighted least squares estimates $(\widehat{S}_\gamma)_{\gamma \in \Gamma}$ for the unknown function $S \in W_r^k$ be given by (3.32). Consider the pair

$$\alpha_0 = (k, t_0)$$

where $t_0 = [\overline{r}/\varepsilon]\varepsilon$, $\overline{r} = r/\varsigma^*$ and ε satisfies the conditions in (3.30). Denote the corresponding weight sequence in Γ as

$$\gamma_0 = \gamma_{\alpha_0} \,. \tag{7.1}$$

Note that for sufficiently large n the pair α_0 belongs to the set (3.29).

Theorem 7.1. *The estimator $\widehat{S}_{\gamma_0}$ satisfies the following asymptotic upper bound*

$$\limsup_{n\to\infty} n^{2k/(2k+1)} \sup_{S \in W_r^k} \mathcal{R}_n^*\,(\widehat{S}_{\gamma_0}, S) \,\leq R_k^* \,. \tag{7.2}$$

Proof. Substituting the model (1.1) in the definition of $\widehat{\theta}_{j,n}$ in (3.4) yields

$$\widehat{\theta}_{j,n} = \theta_j + \frac{1}{\sqrt{n}}\xi_{j,n} \,,$$

where the random variables $\xi_{j,n}$ are defined in (3.4). Therefore, by the definition of the estimators $\widehat{S}_\gamma$ in (3.6), we get

$$\|\widehat{S}_{\gamma_0} - S\|^2 = \sum_{j=1}^n (1-\gamma_0(j))^2\,\theta_j^2 - 2M_n + \frac{1}{n}\sum_{j=1}^n \gamma_0^2(j)\,\xi_{j,n}^2$$

with

$$M_n \,=\, \frac{1}{\sqrt{n}}\sum_{j=1}^n (1\,-\,\gamma_0(j))\,\gamma_0(j)\,\theta_j\,\xi_{j,n} \,.$$

It should be observed that $\mathbf{E}_{Q,S}\,M_n = 0$ for any $Q \in \mathcal{Q}_n^*$. Moreover, by the condition $\mathbf{C}_1$)

$$\mathbf{E}_{Q,S}\sum_{j=1}^n \gamma_0^2(j)\xi_{j,n}^2 \leq \varsigma_Q \sum_{j=1}^n \gamma_0^2(j) + \mathbf{L}_{1,n}(Q)$$

and, taking into account the condition $\mathbf{H}_0$), we get

$$\sup_{Q\in\mathcal{Q}_n} \mathbf{E}_{Q,S} \sum_{j=1}^{n} \gamma_0^2(j)\,\xi_{j,n}^2 \le \varsigma^* \sum_{j=1}^{n} \gamma_0^2(j) + l_n\,.$$

Thus,

$$\mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \,\le\, \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\,\theta_j^2 + \frac{\varsigma^*}{n} \sum_{j=1}^{n} \gamma_0^2(j) + \frac{l_n}{n} \tag{7.3}$$

where $\iota_0 = j_0(\alpha_0)$. Setting

$$\upsilon_n = n^{2k/(2k+1)} \sup_{j\ge\iota_0} (1-\gamma_0(j))^2/a_j\,,$$

we estimate the first summand in the right-hand of (7.3) as

$$n^{2k/(2k+1)} \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\,\theta_j^2 \le \upsilon_n \sum_{j\ge1} a_j\,\theta_j^2\,.$$

¿From here and (6.2), we obtain that for each $S\in W_r^k$

$$\Upsilon_{1,n}(S) =\, n^{2k/(2k+1)} \sum_{j=\iota_0}^{n} (1-\gamma_0(j))^2\,\theta_j^2 \,\le\, \upsilon_n\, r\,.$$

Further we note that

$$\limsup_{n\to\infty}\, (\overline{r})^{2k/(2k+1)}\;\upsilon_n \le\, \frac{1}{\pi^{2k}\,(\tau_k)^{2k/(2k+1)}}\,,$$

where the coefficient τ_k is given in (3.31). Therefore,

$$\limsup_{n\to\infty}\, \sup_{S\in W_r^k} \Upsilon_{1,n}(S) \,\le\, (\varsigma^*)^{2k/(2k+1)}\,\Upsilon_1^* \tag{7.4}$$

where

$$\Upsilon_1^* = \frac{r^{1/(2k+1)}}{\pi^{2k}(\tau_k)^{2k/(2k+1)}}\,.$$

To examine the second summand in the right hand of (7.2), we set

$$\Upsilon_{2,n} = \frac{1}{n^{1/(2k+1)}} \sum_{j=1}^{n} \gamma_0^2(j)\,.$$

It is easy to check that

$$\lim_{n\to\infty} \frac{1}{(\overline{r})^{1/(2k+1)}} \Upsilon_{2,n} = \frac{2(\tau_k)^{1/(2k+1)}\, k^2}{(k+1)(2k+1)} := \Upsilon_2^* .$$

Therefore, taking into account that

$$(\varsigma^*)^{2k/(2k+1)}\, \Upsilon_{1,n}^* + \varsigma^* (\overline{r})^{1/(2k+1)} \Upsilon_2^* = R_k^* ,$$

we obtain

$$\lim_{n\to\infty} n^{2k/(2k+1)} \sup_{S\in W_r^k} \mathcal{R}_n^*(\widehat{S}_{\gamma_0}, S) \le R_k^* .$$

Hence Theorem 7.1. $\square$

7.2 Unknown smoothness

Combining Theorem 7.1 and Theorem 3.6 yields Theorem 6.2.
$\square$

8 Appendix

A.1 Technical lemmas

Lemma A.1. *The operators $\tau_{f,g}$ and $H_{f,g}$ satisfy the following inequalities*

$$\sup_{0\le t\le n} |\tau_{f,g}(t)| \le 4\varpi_{f,g}^* \quad \text{and} \quad \sup_{0\le t\le n} |H_{f,g}(t)| \le \mathbf{D}_1^* \varpi_{f,g}^* , \qquad \text{(A.1)}$$

where $\mathbf{D}_1^*$ is given in (3.19).

Proof. Fist note that

$$\int_0^t f(s)\varepsilon_g(s)\mathrm{d}s = a \int_0^t e^{av} \left(\int_0^{t-v} f(s+v)g(s)(1+e^{2as})\mathrm{d}s \right) \mathrm{d}v .$$

Integrating by parts yields

$$\int_0^{t-v} f(s+v)g(s)(1+e^{2as})\mathrm{d}s = (1+e^{2a(t-v)}) \int_0^{t-v} f(s+v)g(s)\mathrm{d}s$$
$$- 2a \int_0^{t-v} e^{2as} \left(\int_0^s f(z+v)g(z)\mathrm{d}z \right) \mathrm{d}s .$$

Taking into account the definition (4.11), we estimate this integral as

$$\left|\int_0^{t-v} f(s+v)g(s)(1+e^{2as})\mathrm{d}s\right| \le 3\varpi^*_{f,g}\,.$$

Therefore,

$$\left|\int_0^t f(s)\varepsilon_g(s)\mathrm{d}s\right| \le 3\varpi^*_{f,g} \quad \text{and} \quad \left|\int_0^t \varepsilon^*_{f,g}(s)\mathrm{d}s\right| \le 6\varpi^*_{f,g}\,.$$

This implies the first inequality in (A.1). To obtain the second one we represent the function $H_{f,g}(t)$ in the following form

$$\begin{aligned} H_{g,f}(t) &= \lambda\varrho_1^2\tau_{f,g}(t) + \lambda^2\varrho_2^2\int_0^t f(z)g(z)\mathrm{d}z \\ &\quad + \lambda^2\varrho_2^2\left(H^{(1)}_{g,f}(t) + H^{(1)}_{f,g}(t) + H^{(2)}_{g,f}(t) + H^{(2)}_{f,g}(t)\right)\,, \end{aligned}$$

where

$$H^{(1)}_{g,f}(t) = a\int_0^t \int_0^{t-z} e^{ay}g(y+z)f(z)\,\mathrm{d}y\,\mathrm{d}z$$

and

$$H^{(2)}_{g,f}(t) = a^2\int_0^t \int_0^{t-z} e^{ay}g(y+z)\int_0^y e^{av}f(v+z)\,\mathrm{d}v\,\mathrm{d}y\,\mathrm{d}z\,.$$

Now we note

$$|H^{(1)}_{g,f}(t)| = \left|a\int_0^t e^{ay}\left(\int_0^{t-y} g(y+z)f(z)\,\mathrm{d}z\right)\mathrm{d}y\right| \le \varpi^*_{f,g}\,.$$

To estimate $H^{(2)}_{g,f}(t)$ we represent it as

$$H^{(2)}_{g,f}(t) = a^2\int_0^t e^{ay}\left(\int_0^y e^{av}\left(\int_0^{t-y} g(y+z)f(v+z)\mathrm{d}z\right)\mathrm{d}v\right)\mathrm{d}y\,.$$

Note that for any $0 \le v \le y \le t$ one has

$$\left|\int_0^{t-y} g(y+z)f(v+z)\mathrm{d}z\right| \le 2\varpi^*_{f,g}\,.$$

Thus, $|H^{(2)}_{g,f}(t)| \le 2\varpi^*_{f,g}$, and we come to the second inequality in (A.1). Hence Lemma A.1.

□

Lemma A.2. *For any bounded left-continuous* $[0,+\infty) \to \mathbb{R}$ *functions* f, g

$$\mathbf{E}\, v_t(f)\, v_t(g) = \int_0^t e^{2a(t-s)}\, V_{f,g}(s)\, \mathrm{d}s\,,$$

where $V_{f,g}(s)$ *is given in* (4.20).

Proof. By the Ito formula and (4.15), one gets

$$\mathrm{d}\mathbf{E}v_t(f)v_t(g) = 2a\mathbf{E}v_t(f)v_t(g)\mathrm{d}t + (G^*_{f,g}(t) + \varrho_3\, f(t)g(t))\mathrm{d}t$$
$$+ a\,(g(t)\mathbf{E}v_t(f)\,\zeta_t + f(t)\mathbf{E}v_t(g)\,\zeta_t)\,\mathrm{d}t\,.$$

To calculate $\mathbf{E}v_t(f)\,\zeta_t$, we put $g = 1$ in this equality. Then, taking into account that

$$\kappa_f(t) = aG^*_{f,1}(t) + a\varrho_3\, f(t)\,,$$

we get

$$a\mathbf{E}v_t(f)\,\zeta_t = \int_0^t e^{3a(t-s)}\left(f(s)a^2\mathbf{E}\,\zeta_s^2 + \kappa_f(s)\right)\mathrm{d}s = \; A_f(t)\,. \qquad \text{(A.2)}$$

Therefore

$$\mathbf{E}v_t(f)v_t(g) = \int_0^t e^{2a(t-s)}\left(g(s)\,A_f(s) + f(s)\,A_g(s)\right)\mathrm{d}s$$
$$+ \int_0^t e^{2a(t-s)}\left(G^*_{f,g}(s) + \varrho_3\, f(s)g(s)\right)\mathrm{d}s\,.$$

Hence Lemma A.2.

□

Further we will need the following result.

Lemma A.3. *Let* υ *be a continuously differentiable* $\mathbb{R} \to \mathbb{R}$ *function. Then, for any* $n \geq 1$, $\alpha > 0$ *and for any integrated* $\mathbb{R} \to \mathbb{R}$ *function* Ψ,

$$\sup_{0\le t\le n}\left|\int_0^t e^{-\alpha(t-s)}\Psi(s)\upsilon(s)\,\mathrm{d}s\right| \;\le\; \varpi_{1,\Psi}\left(2\|\upsilon\|_* + \frac{\|\dot{\upsilon}\|_*}{\alpha}\right)\,.$$

Proof. One obtains this inequality with the help of integrating by parts.

□

Lemma A.4. *For any mesurable bounded $[0,+\infty)\to\mathbb{R}$ functions f and g, for any $-\infty<a\le 0$, for any $0\le t\le n$ and for any $n\ge 1$*

$$\left|a\int_0^t e^{2a(t-s)}\,g(s)A_f(s)\,\mathrm{d}s\right|\le 3\lambda_2\varpi^*_{f,g}+\varrho_3\|f\|_*\|g\|_*\,. \tag{A.3}$$

Proof. One can represent the function $A_f(t)$ as

$$A_f(t)=\int_0^t e^{3a(t-s)}f(s)\upsilon(s)\mathrm{d}s+\lambda_2\int_0^t e^{3a(t-s)}\varepsilon_f(s)\mathrm{d}s\,, \tag{A.4}$$

where $\upsilon(s)=a^2\mathbf{E}\,\zeta_s^2+\lambda_2\left(e^{2as}-1\right)+a\varrho_3$. ¿From here and (4.23) we rewrite this function as

$$\upsilon(s)=a\upsilon_1(s)+\upsilon_2(s) \tag{A.5}$$

with

$$\upsilon_1(s)=\frac{\varrho_3}{4}\left(e^{4at}+3\right)\quad\text{and}\quad\upsilon_2(s)=\frac{\lambda_2}{2}\,e^{4at}-\frac{\lambda_2}{2}\,.$$

These functions can be estimated as

$$\begin{aligned}\|\upsilon_1\|_* &\le \varrho_3\,;\\ \sup_{-\infty<a\le 0}\left(2\|\upsilon_2\|_*+\frac{\|\dot{\upsilon}_2\|_*}{2|a|}\right) &\le 2\lambda_2\,.\end{aligned} \tag{A.6}$$

Now we represent the intergal in (A.3) as

$$a\int_0^t e^{2a(t-s)}\,g(s)A_f(s)\,\mathrm{d}s=J_1(t)+J_2(t)+\lambda_2J_3(t)\,,$$

where

$$J_1(t)=a^2\int_0^t e^{2a(t-s)}\,g(s)\left(\int_0^s e^{3a(s-u)}f(u)\upsilon_1(u)\mathrm{d}u\right)\,\mathrm{d}s\,,$$

$$J_2(t)=a\int_0^t e^{2a(t-s)}\,g(s)\left(\int_0^s e^{3a(s-u)}f(u)\upsilon_2(u)\mathrm{d}u\right)\,\mathrm{d}s\,,$$

$$J_3(t)=a\int_0^t e^{2a(t-s)}\,g(s)\left(\int_0^s e^{3a(s-u)}\varepsilon_f(u)\mathrm{d}u\right)\,\mathrm{d}s\,.$$

In view of (A.6) we have

$$\sup_{0\le t\le n}|J_1(t)|\le\varrho_3\|f\|_*\|g\|_*\,.$$

Further we represent $J_2(t)$ as

$$J_2(t) = a \int_0^t e^{3au} \left(\int_0^{t-u} e^{2a(t-u-s)} g(s+u) f(s) \upsilon_1(s) \mathrm{d}s \right) \mathrm{d}u \,.$$

By Lemma A.3 and (A.6) we obtain that for any $0 \le z \le n$ and $0 \le u \le n - z$

$$\left| \int_0^z e^{2a(z-s)} \upsilon(s) g(s+u) f(s) \mathrm{d}s \right| \le 3\lambda_2 \varpi^*_{f,g} \,.$$

Therefore,

$$\sup_{0 \le t \le n} |J_2(t)| \le \lambda_2 \, \varpi^*_{f,g} \,.$$

Similarly, one gets

$$\sup_{0 \le t \le n} |J_3(t)| \le \frac{2}{3} \, \varpi^*_{g,\varepsilon_f} \,.$$

To estimate the quantity ϖ_{g,ε_f} defined in (4.11) we note that for any $0 \le v \le n$ and $0 \le t \le n - v$

$$\int_0^t g(s+v) \varepsilon_f(s) \mathrm{d}s = a \int_0^t e^{ax} \, \Theta_{g,f}(t-x, v+x) \mathrm{d}x \,, \tag{A.7}$$

where

$$\Theta_{g,f}(t,v) = \int_0^t g(s+v) f(s) (1 + e^{2as}) \mathrm{d}s \,.$$

Denoting

$$\Upsilon_{g,f}(s,u) = \int_0^s g(r+u) f(r) \mathrm{d}r \,, \tag{A.8}$$

we represent the function $\Theta_{g,f}(t,v)$ as

$$\Theta_{g,f}(t,v) = (1 + e^{2at}) \, \Upsilon_{g,f}(t,v) - 2a \int_0^t e^{2as} \Upsilon_{g,f}(s,v) \mathrm{d}s \,.$$

Therefore

$$\max_{0 \le v \le n} \, \max_{0 \le t \le n-v} |\Theta_{g,f}(t,v)| \, \le \, 3\varpi^*_{f,g} \,.$$

In view of (A.7), one gets

$$\varpi_{g,\varepsilon_f} \le 3\varpi^*_{f,g} \quad \text{and} \quad \sup_{0 \le t \le n} |J_3(t)| \le 2\varpi^*_{f,g} \,.$$

Hence Lemma A.4.

$\square$

Lemma A.5. *For any mesurable bounded* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g*, for any* $-\infty < a \le 0$ *, for any* $0 \le t \le n$ *and for any* $n \ge 1$

$$\left| 2a \int_0^t e^{2a(t-s)}\, G^*_{f,g}(s)\, \mathrm{d}s \right| \le \mathbf{D}^*_2\, \varpi_{f,g}\,, \tag{A.9}$$

where $\mathbf{D}^*_2$ *is defined in* (3.19)*.*

Proof. First we note that the function $G^*_{f,g}$ can be represented as

$$G^*_{f,g}(t) = G_{f,g}(t) + \frac{\lambda_2}{a}\, \dot{\tau}_{f,g}(t) + \frac{\lambda_2}{2a} \left(e^{2at} - 3\right) f(t)g(t)\,.$$

Integrating by the parts yields

$$\left| \int_0^t e^{2a(t-s)}\, \dot{\tau}_{f,g}(s)\, \mathrm{d}s \right| \le 8\varpi^*_{f,g}\,.$$

Finally, by applying Lemma A.3 with $\upsilon(s) = e^{2as} - 3$ and $\Psi(s) = f(s)g(s)$, one gets

$$\left| \int_0^t e^{2a(t-s)}\, (e^{2as} - 3) f(s)g(s)\, \mathrm{d}s \right| \le 7\varpi^*_{f,g}\,.$$

Hence Lemma A.5.

$\square$

Lemma A.6. *For any mesurable bounded* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g*, for any* $-\infty < a \le 0$ *, for any* $0 \le t \le n$ *and for any* $n \ge 1$

$$|\mathcal{V}_{f,g}(t)| \le (6\lambda_2 + \mathbf{D}^*_2)\varpi^*_{f,g} + 3\varrho_3 \|f\|_* \|g\|_*\,. \tag{A.10}$$

Proof. This inequality is a direct consequence of Lemmas A.2-A.5 Hence Lemma A.6.

$\square$

Lemma A.7. *For any mesurable bounded* $[0,+\infty) \to \mathbb{R}$ *functions* f *and* g*, for any* $-\infty < a \le 0$ *,* $0 \le t \le n$ *and* $n \ge 1$

$$\left| a \int_0^t e^{a(t-s)}\, f(s)\mathcal{K}_g(s)\, \mathrm{d}s \right| \le 14\lambda_2 \|g\|_* \varpi^*_{f,g} + \varrho_3 \|f\|_*\, \|g\|^2_*\,. \tag{A.11}$$

Proof. Taking into account (A.4)–(A.5), we write down the function $A_g(t)$ as

$$A_g(t) = A_g^{(1)}(t) + A_g^{(2)}(t)$$

where

$$A_g^{(1)}(t) = a\int_0^t e^{3a(t-s)}g(s)\upsilon_1(s)\mathrm{d}s\,, \quad A_g^{(2)}(t) = \int_0^t e^{3a(t-s)}\left(g(s)\upsilon_2(s) + \lambda_2\varepsilon_g(s)\right)\mathrm{d}s\,.$$

Since

$$U_{1,g}(t) = \frac{2\lambda_2}{a}g(t)\varepsilon_g(t) + \varrho_3 g^2(t)\,,$$

the integral in (A.12) can be represented as

$$a\int_0^t e^{a(t-s)}\,f(s)\mathcal{K}_g(s)\,\mathrm{d}s = J_1^*(t) + J_2^*(t) + J_3^*(t) + J_4^*(t)\,,$$

where

$$J_1^*(t) = 2a^2\int_0^t e^{a(t-s)}f(s)\left(\int_0^s e^{2a(s-r)}g(r)A_g^{(1)}(r)\mathrm{d}r\right)\mathrm{d}s\,,$$

$$J_2^*(t) = 2a^2\int_0^t e^{a(t-s)}f(s)\left(\int_0^s e^{2a(s-r)}g(r)A_g^{(2)}(r)\mathrm{d}r\right)\mathrm{d}s\,,$$

$$J_3^*(t) = 2a\lambda_2\int_0^t e^{a(t-s)}f(s)\left(\int_0^s e^{2a(s-r)}g(r)\varepsilon_g(r)\mathrm{d}r\right)\mathrm{d}s\,,$$

$$J_4^*(t) = \varrho_3\,a^2\int_0^t e^{a(t-s)}f(s)\left(\int_0^s e^{2a(s-r)}g^2(r)\mathrm{d}r\right)\mathrm{d}s\,.$$

In view of (A.6), one obtains

$$\sup_{0\le t\le n}|A_g^{(1)}(t)| \le \frac{\varrho_3}{3}\|g\|_* \quad \text{and} \quad \sup_{0\le t\le n}|J_1^*(t)| \le \frac{\varrho_3}{3}\|f\|_*\,\|g\|_*^2\,.$$

Denoting

$$\Gamma_{f,g}(t,x) = \int_0^t e^{a(t-s)}A_g^{(2)}(s)f(s+x)g(s)\mathrm{d}s\,,$$

one has

$$J_2^*(t) = 2a^2\int_0^{t-x} e^{2ax}\Gamma_{f,g}(t-x,x)\,\mathrm{d}x\,.$$

Noting that

$$\sup_{0\le t\le n} |\upsilon_2(t)| \le \frac{\lambda_2}{2}\,,$$

one comes to the inequalities

$$\sup_{0\le t\le n} |A^{(2)}_g(t)| \le \frac{5\lambda_2}{6|a|}\|g\|_* \quad \text{and} \quad \sup_{0\le t\le n} |\dot{A}^{(2)}_g(t)| \le 4\lambda_2\,\|g\|_*\,.$$

By applying Lemma A.3 with $\Psi(s) = f(s+x)g(s)$ and $\upsilon(s) = \upsilon_2(s)$ one gets

$$\sup_{0\le t\le n}\sup_{0\le x\le t} |\Gamma_{f,g}(t-x,x)| \le \frac{17}{3|a|}\,\lambda_2\,\varpi^*_{f,g}\,\|g\|_* \le 6\lambda_2\,\varpi^*_{f,g}\,\|g\|_*\,.$$

Therefore,

$$\sup_{0\le t\le n} |J^*_2(t)| \le 6\lambda_2\,\varpi^*_{f,g}\,\|g\|_*\,.$$

Similarly, one can show that

$$\sup_{0\le t\le n} |J^*_3(t)| \le 8\,\lambda_2\,\varpi^*_{f,g}\,\|g\|_*\,.$$

Finally, the function $J^*_4(t)$ can be estimated as

$$\sup_{0\le t\le n} |J^*_4(t)| \le \frac{\varrho_3}{2}\|f\|_*\,\|g\|^2_*\,.$$

Hence Lemma A.7.

□

Lemma A.8. *For any measurable bounded* $[0,+\infty)\to\mathbb{R}$ *functions* f *and* g, $-\infty < a \le 0$, $0\le t\le n$ *and* $n\ge 1$

$$\left|a\int_0^t e^{a(t-s)}\,U_{f,g}(s)\,\mathrm{d}s\right| \le 2\mathbf{D}^*_2\,\|g\|_*\varpi^*_{f,g} + \varrho_3\|f\|_*\,\|g\|^2_*\,. \qquad \text{(A.12)}$$

Proof. We note that

$$U_{f,g}(t) = 2g(t)G_{f,g}(t) + \frac{\lambda_2}{a}f(t)g(t)\varepsilon_g(t) + \varrho_3 f(t)g^2(t)\,.$$

Taking into account (A.10) we obtain that

$$\left|2a\int_0^t e^{a(t-s)}\,g(s)G_{f,g}(s)\,\mathrm{d}s\right| \le 2(4\varrho_1^2\varrho^* + \varrho_2\mathbf{D}^*_1)\|g\|_*\varpi^*_{f,g}\,.$$

In view of Lemma A.3, we obtain

$$\sup_{0\le t\le n}\left|\int_0^t e^{a(t-s)}\, f(s)g(s)\varepsilon_g(s)\,\mathrm{d}s\right| \le \varpi^*_{f,g}\left(2\|\varepsilon_g\|_* + \frac{\|\dot{\varepsilon}_g\|_*}{|a|}\right)\,.$$

Taking into account that

$$\|\varepsilon_g\|_* \le 2\|g\|_* \quad \text{and} \quad \|\dot{\varepsilon}_g\|_* \le 4|a|\|g\|_*\,,$$

one gets

$$\sup_{0\le t\le n}\left|\int_0^t e^{a(t-s)}\, f(s)g(s)\varepsilon_g(s)\,\mathrm{d}s\right| \le 8\|g\|_*\varpi^*_{f,g}\,.$$

¿From here we come to desired result. Hence lemma A.8.

□

A.2 Property of the Fourier coefficients

Lemma A.9. *Suppose that the function S in* (1.1) *is differentiable and satisfies the condition* (3.25)*. Then the Fourier coefficients* (3.2) *satisfy the inequality*

$$\sup_{l\ge 2}\, l\sum_{j=l}^{\infty}\theta_j^2 \,\le\, 4\,|\dot{S}|_1^2\,.$$

Proof. In view of (3.1), one has

$$\theta_{2p} = -\frac{1}{\sqrt{2}\pi p}\int_0^1 \dot{S}(t)\sin(2\pi p t)\mathrm{d}t$$

and

$$\begin{aligned}\theta_{2p+1} &= \frac{1}{\sqrt{2}\pi p}\int_0^1 \dot{S}(t)(\cos(2\pi p t)-1)\mathrm{d}t\\ &= -\frac{\sqrt{2}}{\pi p}\int_0^1 \dot{S}(t)\sin^2(\pi p t)\mathrm{d}t\,, \quad p\ge 1\,.\end{aligned}$$

¿From here, it follows that for any $j\ge 2$

$$\theta_j^2 \,\le\, \frac{2}{j^2}|\dot{S}|_1^2\,.$$

Taking into account that

$$\sup_{l\ge 2} l\sum_{j\ge l}\frac{1}{j^2} \,\le\, 2\,,$$

we arrive at the desired result. □

9 Acknowledgments

The paper is supported by the RFBI-Grant 09-01-00172-a and the Russian State Contract 02.740.11.5026.

References

[1] Akaike, H.: A new look at the statistical model identification. *IEEE Trans. on Automatic Control*, **19**, 716–723 (1974).

[2] Barndorff-Nielsen, O.E. and Shephard, N.: Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial mathematics. *J. Royal Stat. Soc.*, **B 63**, 167–241 (2001).

[3] Barron, A., Birgé, L. and Massart, P.: Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, **113**, 301–415 (1999).

[4] Bertoin, J.: *Lévy Processes*. Cambridge University Press, Cambridge (1996).

[5] Brua, J.: Asymptotically efficient estimators for nonparametric heteroscedastic regression models. *Stat. Methodol.*, **6**(1), 47–60 (2009).

[6] Delong, L. and Klüppelberg, C.: Optimal investment and consumption in a Black-Scholes market with Lévy driven stochastic coefficients. *Annals of Applied Probability*, **18**(3), 879-908 (2008).

[7] Fourdrinier, D. and Pergamenshchikov, S.: Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*, **59**(3), 435–464 (2007).

[8] Ferraty, F. and Vieu, P.: *Nonparametric Functional Data Analysis : Theory and Practice*. Springer Series in Statistics, Springer-Verlag, New York (2006).

[9] Galtchouk, L. and Pergamenshchikov, S.: Nonparametric sequential estimation of the drift in diffusion processes. *Mathematical Methods of Statistics*, **13**(1), 25–49 (2004).

[10] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotically efficient estimates for non parametric regression models.*Statistics and Probability Letters*, **76**(8), 852–860 (2006).

[11] Galtchouk, L. and Pergamenshchikov, S.: Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21**(1), 1–16 (2009).

[12] Galtchouk, L. and Pergamenshchikov, S.: Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression.*Journal of Korean Statistical Society*, **38**(4), 305–322 (2009).

[13] Galtchouk, L. and Pergamenshchikov, S.: Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression via model selection, *http://hal.archives-ouvertes.fr/hal-00326910/fr/*, (2009).

[14] Goldfeld, S.M. and Quandt, R.E.: *Nonlinear Methods in Econometrics*. North-Holland, London (1972).

[15] Kneip, A. Ordered linear smoothers. *Annals of Statistcs*, **22**, 835–866 (1994).

[16] Konev, V.V. and Pergamenshchikov, S.M.: Sequential estimation of the parameters in a trigonometric regression model with the gaussian coloured noise. *Statistical Inference for Stochastic Processes*, **6**, 215–235 (2003).

[17] Konev, V.V. and Pergamenshchikov, S.M.: General model selection estimation of a periodic regression with a Gaussian noise. - *Annals of the Institute of Statistical Mathematics*, (2010). *http://dx.doi.org/10.1007/s10463-008-0193-1*

[18] Konev, V.V. and Pergamenshchikov, S.M.: Nonparametric estimation in a semimartingale regression model. Part 1. Oracle Inequalities. *Vestnik Tomskogo Universiteta, Mathematics and Mechanics*, **3**(7), 23–41 (2009). *http://hal.archives-ouvertes.fr/hal00417603/fr*

[19] Konev, V.V. and Pergamenshchikov, S.M.: Nonparametric estimation in a semimartingale regression model. Part 2. Robust asymptotic efficiency. *Vestnik Tomskogo Universiteta, Mathematics and Mechanics*, **4**(8), 31–45 (2009). *http://hal.archives-ouvertes.fr/hal-00417600/fr*

[20] Jacod, J. and Shiryaev, A.N.: *Limit theorems for stochastic processes*. Vol.1, Springer, New York (1987).

[21] Mallows, C.:Some comments on C_p. *Technometrics*, **15**, 661–675 (1973).

[22] Massart, P.: A non-asymptotic theory for model selection. *4ECM Stockholm*, 309-323 (2004).

[23] Nussbaum, M.: Spline smoothing in regression models and asymptotic efficiency in $\mathbf{L}_2$. *Ann. Statist.*, **13**, 984-997 (1985).

[24] Pinsker, M.S.: Optimal filtration of square integrable signals in gaussian white noise. *Problems of Transimission information*, **17**, 120–133 (1981).